Learning Human Behavior From Data (Part 1)

Lecture #8
Recap of last 3 lectures

• Sentiment
• Attitudes
• Beliefs

• Today: Behaviors
Social Phenomena

• When people interact, the language reflects social phenomena
  – We want to understand and computationally model these phenomena
Examples of social phenomena

• Social roles
• Behaviors
• Intentions
• Structure
• Hierarchy
• Any changes to any of the above
Behaviors

• We are primarily interested in socio-linguistic behaviors
  – There are behaviors that are social in nature
  – And can be expressed through language
Examples of sociolinguistic behaviors

• Topic Control

• others?
Topic Control

• *Topic Control*: attempts by a discourse participant to impact or impose the topic of conversation.

  – Can be observed over a length of discourse
  – Relative to other participants’ behavior.
  – Speakers display varying degrees of Topic Control
Topic Control indices

• What can we detect from language that might be a component of topic control behavior of a person?
Topic Control indices

- Topic Introductions (LTI)
- Subsequent Mentions (SMT)
- Citation Score (CS)
- Turn Length (TL)
Topic Control indices

• Local Topic Introductions (LTI)
  – Speakers who introduce more local topics, exert more topic control in dialogue. The LTI index counts the number of local topics introduced by each participant as percentage of all local topics in a discourse.
  – Local topics are defined as noun phrases introduced into discourse, which are subsequently mentioned again via repetition, synonym, or pronoun.
Turn 42- **LE (8:33:03 PM)**: I guess we should just start, not wait for CS and SH?

Turn 43- **JR (8:33:32 PM)**: sure

Turn 44- **KN (8:33:43 PM)**: ok

Turn 45- **LE (8:34:02 PM)**: Fundraising was Mark, Nanny was **Carla**, I think, if you were talking about my comment.

Turn 46- **JR (8:35:05 PM)**: gotcha- so that is not he most important to get this job…. 

Turn 47- **JR (8:35:23 PM)**: sorry about my typos- not used to this laptop yet

Turn 48- **JR (8:36:27 PM)**: wanna go thru carlas resume first ?

Turn 49- **KN (8:36:43 PM)**: sure

Turn 50- **LE (8:36:44 PM)**: Sure.

Turn 51- **KN (8:37:00 PM)**: i wonder how old carla is

Turn 52- **LE (8:37:24 PM)**: Ha, yeah, when I hear nanny I think someone older.

Turn 53- **KN (8:37:30 PM)**: she's got a perfect driving record and rides **horses**! coincidence?

Turn 54- **JR (8:37:35 PM)**: '06 high school grad

Turn 55- **KN (8:37:44 PM)**: i think she rides a **horse** and not a car!
Topic introductions → Topic control

Turn 42- **LE (8:33:03 PM)**: I guess we should just start, not wait for CS and SH?

Turn 43- **JR (8:33:32 PM)**: sure

Turn 44- **KN (8:33:43 PM)**: ok

Turn 45- **LE (8:34:02 PM)**: Fundraising was Mark, Nanny was Carla, I think, if you were talking about my comment.

Turn 46- **JR (8:35:05 PM)**: gotcha- so that is not he most important to get this job....

Turn 47- **JR (8:35:23 PM)**: sorry about my typos- not used to this laptop yet

Turn 48- **JR (8:36:27 PM)**: wanna go thru Carla’s resume first?

Turn 49- **KN (8:36:43 PM)**: sure

Turn 50- **LE (8:36:44 PM)**: Sure.

Turn 51- **KN (8:37:00 PM)**: i wonder how old Carla is

Turn 52- **LE (8:37:24 PM)**: Ha, yeah, when I hear nanny I think someone older.

Turn 53- **KN (8:37:30 PM)**: she's got a perfect driving record and rides horses! coincidence?

Turn 54- **JR (8:37:35 PM)**: '06 high school grad

Turn 55- **KN (8:37:44 PM)**: i think she rides a horse and not a car!
Topic Control indices

• Turn Length (TL)
  – Speakers who have, on average, longer turns exert more topic control in discourse. Turn length reflects (a) the extent to which participants attempt to control the topic of subsequent discussion; and (b) the extent to which other participants are willing to ‘yield the floor’ to various participants.
  – The TL index is the average number of words per turn for each speaker.
Turn length →

Topic control

- Turn 42- LE (8:33:03 PM): I guess we should just start, not wait for CS and SH?
  - Turn 43- JR (8:33:32 PM): sure
  - Turn 44- KN (8:33:43 PM): ok

Turn 45- LE (8:34:02 PM): Fundraising was Mark, Nanny was Carla, I think, if you were talking about my comment.
  - Turn 46- JR (8:35:05 PM): gotcha- so that is not he most important to get this job….
  - Turn 47- JR (8:35:23 PM): sorry about my typos- not used to this laptop yet
  - Turn 48- JR (8:36:27 PM): wanna go thru carlas resume first?
  - Turn 49- KN (8:36:43 PM): sure

Turn 50- LE (8:36:44 PM): Sure.
  - Turn 51- KN (8:37:00 PM): i wonder how old carla is

Turn 52- LE (8:37:24 PM): Ha, yeah, when I hear nanny I think someone older.
  - Turn 53- KN (8:37:30 PM): she's got a perfect driving record and rides horses! coincidence?
  - Turn 54- JR (8:37:35 PM): '06 high school grad
  - Turn 55- KN (8:37:44 PM): i think she rides a horse and not a car!
Turn length ➔

Topic control

- Turn 42- LE (8:33:03 PM): I guess we should just start, not wait for CS and SH?
  Turn 43- JR (8:33:32 PM): sure
  Turn 44- KN (8:33:43 PM): ok
  Turn 45- LE (8:34:02 PM): Fundraising was Mark, Nanny was Carla, I think, if you were talking about my comment.
  Turn 46- JR (8:35:05 PM): gotcha- so that is not he most important to get this job…. 
  Turn 47- JR (8:35:23 PM): sorry about my typos- not used to this laptop yet
  Turn 48- JR (8:36:27 PM): wanna go thru carlas resume first ?
  Turn 49- KN (8:36:43 PM): sure
  Turn 50- LE (8:36:44 PM): Sure.
  Turn 51- KN (8:37:00 PM): i wonder how old carla is
  Turn 52- LE (8:37:24 PM): Ha, yeah, when I hear nanny I think someone older.
  Turn 53- KN (8:37:30 PM): she's got a perfect driving record and rides horses! coincidence?
  Turn 54- JR (8:37:35 PM): '06 high school grad
  Turn 55- KN (8:37:44 PM): i think she rides a horse and not a car!
# The result:
Topic Control map for a discourse

<table>
<thead>
<tr>
<th>speaker</th>
<th>Topic introductions</th>
<th>Topic mentions</th>
<th>Citing others</th>
<th>Turn length</th>
<th>Degree of Topic Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>JR</td>
<td>4</td>
<td>0.21</td>
<td>5</td>
<td>0.28</td>
<td>5</td>
</tr>
<tr>
<td>KI</td>
<td>3</td>
<td>0.19</td>
<td>3</td>
<td>0.20</td>
<td>4</td>
</tr>
<tr>
<td>KA</td>
<td>2</td>
<td>0.10</td>
<td>2</td>
<td>0.10</td>
<td>1</td>
</tr>
<tr>
<td>LE</td>
<td>5</td>
<td>0.27</td>
<td>4</td>
<td>0.22</td>
<td>3</td>
</tr>
<tr>
<td>KN</td>
<td>3</td>
<td>0.11</td>
<td>3</td>
<td>0.11</td>
<td>3</td>
</tr>
<tr>
<td>CS</td>
<td>1</td>
<td>0.09</td>
<td>1</td>
<td>0.06</td>
<td>2</td>
</tr>
</tbody>
</table>
Topic Control indices correlations
Let’s look at another behavior

• *Involvement* is defined as a degree of engagement or participation in the discussion of a group. A degree of involvement may be estimated by how much a speaker contributes to the discourse in terms of substantive content. Highly involved speakers are particularly engaged in discussion of the most important and persistent topics.
Involvement indices

- Involvement
- Allotopicality (ATP)
- All Subsequent Mentions (ASM)
- Noun Phrase Index (NP)
- Turn Index (TI)
- Topic Chain Index (TCI)
Involvement indices – Turn Index

• Turn Index (TI)
  – Counts the ratio of turns per participant to the total number of turns in the discourse.
Involvement indices – Topic Chain Index

• The *Topic Chain Index (TCI)* counts the degree to which participants are involved in discussion of the most persistent topics.

• In order to calculate TCI values, we define a *topic chain* as a noun phrase and all subsequent mentions of it (whether these subsequent mentions involve repetition, a synonym, or a pronoun). The longest of these chains indicate the most persistent topics. We compute the percentages of mentions of these topics for each participant.
Topic Chain Index ➔
Involvement

- Turn 42- **LE (8:33:03 PM)**: I guess we should just start, not wait for CS and SH?
  - Turn 43- **JR (8:33:32 PM)**: sure
  - Turn 44- **KN (8:33:43 PM)**: ok
- Turn 45- **LE (8:34:02 PM)**: Fundraising was Mark, Nanny was Carla, I think, if you were talking about my comment.
- Turn 46- **JR (8:35:05 PM)**: gotcha- so that is not he most important to get this job....
- Turn 47- **JR (8:35:23 PM)**: sorry about my typos- not used to this laptop yet
- Turn 48- **JR (8:36:27 PM)**: wanna go thru Carla's resume first?
- Turn 49- **KN (8:36:43 PM)**: sure
- Turn 50- **LE (8:36:44 PM)**: Sure.
- Turn 51- **KN (8:37:00 PM)**: i wonder how old Carla is
- Turn 52- **LE (8:37:24 PM)**: Ha, yeah, when I hear nanny I think someone older.
- Turn 53- **KN (8:37:30 PM)**: she's got a perfect driving record and rides horses! coincidence?
- Turn 54- **JR (8:37:35 PM)**: '06 high school grad
- Turn 55- **KN (8:37:44 PM)**: i think she rides a horse and not a car!
Topic chains → Involvement

- Turn 42- **LE (8:33:03 PM)**: I guess we should just start, not wait for CS and SH?
- Turn 43- **JR (8:33:32 PM)**: sure
- Turn 44- **KN (8:33:43 PM)**: ok
- Turn 45- **LE (8:34:02 PM)**: Fundraising was Mark, Nanny was **Carla**, I think, if you were talking about my comment.
- Turn 46- **JR (8:35:05 PM)**: gotcha- so that is not he most important to get this job....
- Turn 47- **JR (8:35:23 PM)**: sorry about my typos- not used to this laptop yet
- Turn 48- **JR (8:36:27 PM)**: wanna go thru **carlas** resume first ?
- Turn 49- **KN (8:36:43 PM)**: sure
- Turn 50- **LE (8:36:44 PM)**: Sure.
- Turn 51- **KN (8:37:00 PM)**: i wonder how old **carla** is
- Turn 52- **LE (8:37:24 PM)**: Ha, yeah, when I hear nanny I think someone older.
- Turn 53- **KN (8:37:30 PM)**: **she**'s got a perfect driving record and rides **horses**! coincidence?
- Turn 54- **JR (8:37:35 PM)**: '06 high school grad
- Turn 55- **KN (8:37:44 PM)**: i think **she** rides a **horse** and not a car!
Let’s look at yet another behavior

- **Task Control** is an effort by one or more members of the group to define the group’s project or goal and/or steer the group towards that goal.

- Usually done via directive statements

- A directive is successful if it achieves the desired effect, which may include changing the topic of conversation or obtaining assents from other participants. A person who uses more directives and whose directives are successful more often, is considered to have a high degree of Task Control in conversation.
Task Control indices

- Directive Index (DI)
- Directed Topic Shift Index (DTSI)
- Process Management Index (PMI)
- Process Management Success Index (PMSI)
Task Control indices – Directive Index

• Directive Index (DI): A participant who directs others is attempting to control the task that the group is performing. Similarly, a participant who offers and commits to perform a task is attempting to control the task. We count the number of directives and offers, i.e., utterances where one speaker asks or directs another, or offers/commits self to do something, made by each participant as a percentage of all directives and offers in discourse.
Turn 42- LE (8:33:03 PM): I guess we should just start, not wait for CS and SH?
Turn 43- JR (8:33:32 PM): sure
Turn 44- KN (8:33:43 PM): ok
Turn 45- LE (8:34:02 PM): Fundraising was Mark, Nanny was Carla, I think, if you were talking about my comment.
Turn 46- JR (8:35:05 PM): gotcha- so that is not he most important to get this job….
Turn 47- JR (8:35:23 PM): sorry about my typos- not used to this laptop yet

Turn 48- JR (8:36:27 PM): wanna go thru carlas resume first ?
Turn 49- KN (8:36:43 PM): sure
Turn 50- LE (8:36:44 PM): Sure.
Turn 51- KN (8:37:00 PM): i wonder how old carla is
Turn 52- LE (8:37:24 PM): Ha, yeah, when I hear nanny I think someone older.
Turn 53- KN (8:37:30 PM): she's got a perfect driving record and rides horses! coincidence?
Turn 54- JR (8:37:35 PM): '06 high school grad
Turn 55- KN (8:37:44 PM): i think she rides a horse and not a car!
Task Control indices – Process Management Success Index

• PMSI: A speaker whose suggestions, directives, and offers are met with consents of other participants exerts a degree of task control in discourse.

• To determine the degree of task control by each speaker, we give one credit for each process management suggestion, directive, and offer made by the speaker, and one credit for each supporting response by another speaker. Conversely, we take away a credit for each response that rejects or qualifies such utterances.
Process Management Success Index - Task Control

Turn 42- LE (8:33:03 PM): I guess we should just start, not wait for CS and SH?
   Turn 43- JR (8:33:32 PM): sure
   Turn 44- KN (8:33:43 PM): ok
   Turn 45- LE (8:34:02 PM): Fundraising was Mark, Nanny was Carla, I think, if you were talking about my comment.
   Turn 46- JR (8:35:05 PM): gotcha- so that is not he most important to get this job.…. 
   Turn 47- JR (8:35:23 PM): sorry about my typos- not used to this laptop yet

**Turn 48- JR (8:36:27 PM): wanna go thru carlas resume first ?**
   Turn 49- KN (8:36:43 PM): sure
   Turn 50- LE (8:36:44 PM): Sure,
   Turn 51- KN (8:37:00 PM): i wonder how old carla is
   Turn 52- LE (8:37:24 PM): Ha, yeah, when I hear nanny I think someone older.
   Turn 53- KN (8:37:30 PM): she's got a perfect driving record and rides horses! coincidence?
   Turn 54- JR (8:37:35 PM): '06 high school grad
   Turn 55- KN (8:37:44 PM): i think she rides a horse and not a car!
How to compute the indices automatically?

• Local Topic Index
How to compute the indices automatically?

• Local Topic Index
  – Need a Part-of-Speech Tagger
  – Need a co-reference resolver
How to compute the indices automatically?

• Local Topic Index
  – Need a Part-of-Speech Tagger
  – Need a co-reference resolver

• Turn Length
• Turn Index
How to compute the indices automatically?

• Local Topic Index
  – Need a Part-of-Speech Tagger
  – Need a co-reference resolver

• Turn Length

• Turn Index

• Topic Chain Index
  – ?
How to compute the indices automatically?

- Local Topic Index
  - Need a Part-of-Speech Tagger
  - Need a co-reference resolver

- Turn Length
- Turn Index

- Topic Chain Index
  - Need a Part-of-Speech Tagger
  - Need a co-reference resolver
How to compute the indices automatically?

- Directive Index
How to compute the indices automatically?

• Directive Index
  – Need a dialogue act tagger
  – What is dialogue act tagger?
  – Give a handle to the functional meaning of an utterance
    • *E.g. Can you close the door?*
How to compute the indices automatically?

• Directive Index
  – Need a dialogue act tagger
  
  – What is dialogue act tagger?
  – Give a handle to the functional meaning of an utterance

  • E.g. Can you close the door?
    – is not a question!
How to compute the indices automatically?

• Directive Index
  – Need a dialogue act tagger

• Process Management Success Index
  – Need a dialogue act tagger
  – Need a communicative act tagger
How to compute the indices automatically?

• Directive Index
  – Need a dialogue act tagger

• Process Management Success Index
  – Need a dialogue act tagger
  – Need a communicative act tagger
    • A communicative act tagger automatically ascertains whether an utterance is 1) addressed to someone; 2) a response to someone; or 3) a continuation of a prior utterance
Modeling sociolinguistic behaviors

• Using a few simple NLP tools, we can compute complex behaviors
• We can rank participants on the relative degree to which they exhibit these behaviors